Identifiers in EAC-CPF
August 1, 2019
Katherine M. Wisser

The generic definition of identifier according to many metadata standards is “an unambiguous reference in a local context.”

Types of identifiers

There are three types of identifiers that need to be handled vis-à-vis the various tasks performed within an EAC-CPF record. The use cases below seek to illustrate and make distinct these three categories.

Type 1: Database primary key

Definition: A primary key is a special relational database table column (or combination of columns) designated to uniquely identify all table records. A primary key's main features are: It must contain a unique value for each row of data. It cannot contain null values.

The concept of a primary key is borrowed from the relational database environment to assign individual *records* in a given system (context) a unique identifier. These identifiers are typically system-assigned and have no content connection to the record that they represent. They primarily distinguish one record from another record in a data set (in a database table it would uniquely distinguish one row from another row).

Type 2: Identifiers that distinguish and determine entities

This category breaks down into two types: informational (contains information that is meaningful in terms of identifying the entity) and non-informational (similar to a primary key, this is system-defined and while built through a structure does not necessarily build on contextual information about the entity).

<u>Informational</u>
Examples of the first type include alphanumeric strings that represent names, including persons, corporate bodies, geographic places, meetings, uniform titles, and to a lesser extent family names. In this first group, appropriate authority work is done to establish the string, account for variants, and provide contextual information to work in conjunction with the string to disambiguate that entity from others that may have the same or similar string.

Examples of variations on names From ISNI

Alternative spellings: William Shakespear, William Shakespeare
Alternative presentations: e.e. cummings, E.E. Cummings
Character set variances: Günter Graβ, Guenter Grass, and Guenter Graβ
Transliteration variances: Ciaikovsky, Pjotr Iijc and Пётр Ильич Чайковский
Linguistic variances: Pyotr Tchaikovsky, Peter Tchaikovsky

<u>Non-informational</u>

Examples of non-informational identifiers of this type are alphanumeric (primarily numeric in the systems that have been identified) strings that do not have a meaningful connection to the entity that they represent. Examples include ISNI and ORCID

ORCID: a persistent digital identifier, researcher focus, https://orcid.org

> Format of the ORCID iD
> The ORCID iD is an https URI with a 16-digit number that is compatible with the ISO Standard (ISO 27729), also known as the International Standard Name Identifier (ISNI), e.g. https://orcid.org/0000-0001-2345-6789
>
> Initially ORCID iDs will be randomly assigned by the ORCID Registry from a block of numbers that will not conflict with ISNI-formatted numbers assigned in other ways. ORCID iDs always require all 16 digits of the identifier; they can not be shortened to remove leading zeros if they exist.
>
> No information about a person is encoded in the ORCID iD. The identifiers were designed to be usable in situations where personally-identifiable information should/can not be shared. Also, since the ORCID iD is designed to be a career-long identifier, no information that can change over a person's career is embedded in the iD, e.g., country, institution, field of study.
>
> Issuing ORCID iDs
> Only the ORCID Registry can assign ORCID iDs. ORCID iDs are intended to be assigned to individuals and may be secured at no charge. ORCID iDs will be assigned randomly from a block of numbers reserved for this purpose. Typically, the 16-digit identifiers are assigned between 0000-0001-5000-0007 and 0000-0003-5000-0001.

ISNI: International Standard Name Identifier (ISO 27729), http://www/isni.org
Standard number of contributors to creative works and distribution, includes public name and a persistent unique identifying number to resolve problems of name ambiguity in search and discovery; intended to be a bridge identifier across multiple domains.

Type 3: *Identifiers that create unique locations within an XML instance*

A unique identifier that establishes a distinct place internally within an XML instance. The content of the identifier has no informational value aside from its uniqueness within the instance that facilitates identification of that location in an XML instance.

Use cases

- Establishing an identifier for the XML instance (record) being created (TYPE 1)
- Maintaining information about identifiers previously assigned to the XML instance (TYPE 1)
- Identifiers assigned to records in other systems (TYPE 1)
- Identifiers used as a representation of the entity being described in a record (TYPE 2)
- Data values used as a representation of something and articulated in a data value standard (TYPE 2)
- Identifiers established within an XML instance for distinction and navigation (TYPE 3)

Sections of the EAC-CPF schema in which the notion of identifier is relevant

recordId
A required element within <control> that designates a unique identifier for the EAC-CPF instance. The assigning owner ensures the uniqueness of the <recordId> within the EAC-CPF descriptive system under its control. The <recordId>, when used in combination with the content of the required <agencyCode> element within <maintenanceAgency>, will provide a globally unique identifier.

Record alternate record identifiers if desired in <otherRecordId>.

otherRecordId
An element in <control> used to encode record identifiers that are alternatives to the mandatory identifier in <recordId>. These might include the identifiers of merged EAC-CPF instances representing the same CPF entity or those of records that are no longer current but had some part in the history and maintenance of the EAC-CPF instance.

The localType attribute can be used to identify the kind of institution or service responsible for each associated record identifier if not the same as that given in the <maintenanceAgency> element for this EAC-CPF instance.

source
An element for identifying a particular source of evidence used in describing the CPF entity. A record for the source must be included in either <objectXMLWrap> or <objectBinWrap>; or as a textual description in the <sourceEntry> element. Use the optional <descriptiveNote> for any additional notes about the source. A <source> in this context should not be confused with the <citation> element which is used in a number of descriptive elements to point to a resource that provides descriptive data which is not otherwise given in the EAC-CPF instance.

A link to the source may be made using the XML Linking Language (Xlink) attributes (consult the specification at http://www.w3.org/TR/xlink/) and the last date and time that the source was verified can be given in the lastDateTimeVerified attribute.

entityId
An optional element of <identity> that may be used to record any identifier associated with the CPF entity being described in the EAC-CPF instance. Identifiers such as legal identifiers, typically assigned by an authoritative agency, may be recorded in this element.

Do not confuse with <recordId> within <control>, which refers to an identifier for the EAC-CPF instance rather than the entity it describes.

nameEntry
Within <identity>, the element <nameEntry> is used to record a name by which the corporate body, the person, or the family described in the EAC-CPF instance is known.

When <nameEntry> occurs within <nameEntryParallel> it is used to record two or more parallel forms (e.g., official forms of the name in different languages and/or scripts or transliterated forms of the name). When <nameEntry> is not included within <nameEntryParallel> it is used to record the authorized or alternative forms, whether standardized or not.

Each form of the name is recorded in a separate <nameEntry> element.

Each <nameEntry> should contain at least one <part> element. Within <nameEntry> each of the component parts of a name may be recorded in a separate <part> element.

When <nameEntry> does not occur within <nameEntryParallel>, it may include two optional elements, <authorizedForm> and <alternativeForm>, to account more precisely for the status of the form of the name contained in the <nameEntry> element, as compared to other possible forms of the name contained in other <nameEntry> elements.

The <nameEntry> element may also contain a <useDates> element to indicate the dates the name was used but only when <nameEntry> is not included within <nameEntryParallel>.
The attributes scriptCode and xml:lang are used to specify the script and the language of each of the names recorded in <nameEntry>.

If the form of the name recorded in <nameEntry> is a transliterated one, the attribute transliteration is used to record the conventions or rules applied to transliterate this form of the name.

cpfRelation cpfRelationType="identity"
The <cpfRelation> element contains the description of a corporate body, person, or family related to the described CPF entity. Such related entities are typically described in another EAC-CPF instance or other encoding language such as MARC. Use the <objectXMLWrap> to incorporate XML elements from any XML namespace or <objectBinWrap> for base64-encoded binary data. A <relationEntry> element is provided for textual identification of the related entity.

Use the <date>, <dateRange>, or <dateSet> elements for specifying the time period of the relationship and the <placeEntry> element for recording relevant location information. A <descriptiveNote> element may be included for a more detailed explanation of the relationship.

The cpfRelationType attribute may be used to specify the nature of the <cpfRelation> entity's relationship to the entity described in the EAC-CPF instance. Values are chosen from a closed list.

@cpfRelationType="identity" is an equivalence relationship.

@xlm:id
An identifier used to name the element so that it can be referred to, or referenced from, somewhere else. Each xml:id within a document must have a unique value. The xml:id attribute regularizes the naming of the element and thus facilitates building links between it and other resources. For example, the xml:id may be used to uniquely identify two or more <cpfDescription> within <multipleIdentities>. Uniquely identifying or distinguishing two or more <cpfDescription> may be essential or useful in maintenance environments when relating resources, functions, or corporate bodies, persons, or families to one among two or more identities represented in one EAC-CPF instance.

Identifiers in other metadata schema

Outlined below are identifiers of various types from other related metadata schemas. These standards include Dublin Core, EAD3, MARC, METS, MODS, and TEI.

Dublin Core

Identifier
Definition: An unambiguous reference to the resource within a given context.
Comment: Recommended best practice is to identify the resources by means of a string conforming to a formal identification system.

(TYPE 1)

Encoding Scheme. These qualifiers identify schemes that aid in the interpretation of an element value. These schemes include controlled vocabularies and formal notations or parsing rules. A value expressed using an encoding scheme will thus be a token selected from a controlled vocabulary (e.g., a term from a classification system or set of subject headings) or a string formatted in accordance with a formal notation (e.g., "2000-01-01" as the standard expression of a date). If an encoding scheme is not understood by a client or agent, the value may still be useful to a human reader. The definitive description of an encoding scheme for qualifiers must be clearly identified and available for public use.

Summary of DC elements with specified/recommended encoding schemes:

DCMES Element	Element Encoding Scheme(s)
Subject	LCSH MeSH DDC LCC UDC
Date	DCMI Period W3C-DTF
Type	DCMI Type Vocabulary
Format (medium)	IMT
Identifier	URI
Source	URI
Language	ISO 639-2 RFC 1766
Relation	URI
Coverage (spatial)	DCMI Point ISO 3166 DCMI Box

	TGN
Coverage (temporal)	DCMI Period W3C-DTF

EAD3

<recordid>

A required child element of <control> that designates a unique identifier for the EAD instance.

(TYPE 1)

<otherrecordid>

A child element of <control> that encodes any local identifier for the EAD instance.

(TYPE 1)

Names

<persname>

An element for identifying a personal name.

<corpname>

An element for identifying the name of an organization or group of people.

<name>

An element for encoding generic names.

<geogname>

An element for encoding place names.

Use @identifier to provide a number, code, or string (e.g., URI) that uniquely identifies the personal name in a controlled vocabulary, taxonomy, ontology, or other knowledge organization system. Do not confuse with @id, which provides a unique id for the element within the XML instance.

(TYPE 2: Informational)

@id

An identifier that must be unique within the current document and is used to name the element so that it can be referred to, or referenced from, somewhere else. This facilitates building links between the element and other resources. Do not confuse with identifier, which provides a machine-processable identifier for an entity or concept in an external system.

Data Type:

ID

(TYPE 3)

MARC

In a MARC authority record, a **heading** is the content of a 1XX, 4XX, or 5XX field that documents the form of heading used for indexing and retrieval or organizational purposes in a file. Two types of headings are defined in the authorities format:

- **Established heading** - A heading that is authorized for use in other MARC records as a main entry (1XX), added entry (700-730), or series added entry (440 or 800-830) field or as the lead element in a subject access (600-655; 654-657) field. In authority records, established headings are used in fields 100-155 (headings) and fields 500-555 (tracings) for established heading (008/09, Kind of record, code a or f) records.
- **Unestablished heading** - A heading that is *not* authorized for use in other MARC records as the lead element of a main, added, series, or subject access field. An unestablished heading may be a reference to a variant form of the established heading, a form of the heading used only for authority file organizational purposes, or a subject subdivision that is authorized for use with an established heading in an extended subject heading. In authority records, unestablished headings are used in the 1XX (heading) and 4XX (tracing) fields of reference (008/09, code b or c), subdivision (code d), reference and subdivision (code g), and node label (code e) records. An unestablished heading may also be used in the 4XX fields of established heading (code a or f) records.

(TYPE 2: Informational)

The Library of Congress Card Number was the number used to identify and control catalog cards. With the development of the MARC format and the first distribution of machine-readable records for book materials in the late 1960s, the name of the LCCN was changed to Library of Congress Control Number. LCCNs are used for authority, bibliographic and classification records and are currently structured as follows:

Element	Length	Positions
Alphabetic Prefix	3	00-02
Year	2	03-04
Serial Number	6	05-10
Supplement Number	1	11

The uniqueness of the LCCN is determined by the first 11 positions (positions 00-10). The Supplement Number has never been used by the Library of Congress and this position is always blank. The Supplement Number may be followed by two kinds of variable length data known as Suffix/Alphabetic Identifier and Revision Date. Each Suffix/Alphabetic Identifier is preceded by a slash as is Revision Date. If there is no Suffix/Alphabetic Identifier, the Revision Date is preceded by two slashes. Examples:

###95156543# May be displayed as: 95-156543
###94014580#/AC/r95 May be displayed as: 94-14580/AC/r95
###79310919#//r86 May be displayed as: 79-310919//r86
gm#71005810# May be displayed as: gm71-5810

For more details on the current and future structure of the Library of Congress Control Number, see the Structure of the Library of Congress Control Number.

MARC Authority 001
FIELD DEFINITION AND SCOPE
Control number assigned by the organization creating, using, or distributing the record. For interchange purposes, documentation of the structure of the control number and input conventions should be provided to exchange partners by the organization initiating the interchange. The MARC code identifying whose system control number is present in field 001 is contained in field 003 (Control Number Identifier). An organization using a record of another organization may move the incoming control number from field 001 (and the control number identifier from field 003) to field 035 (System Control Number), 010 (Library of Congress Control Number), or 016 (National Bibliographic Agency Control Number), as appropriate, and place its own system control number in field 001 (and its control number identifier in field 003).

MARC Authority 010
FIELD DEFINITION AND SCOPE
Unique number assigned to a record by the Library of Congress (LC) or a cooperative cataloging partner contributing authority records to the Name Authority Cooperative Program (NACO) database. The field is also assigned to records created by LC for the Library of Congress Subject Headings (LCSH).

LC control number is carried in field 010 subfield $a and field 001 (Control Number) in records distributed by LC's Cataloging Distribution Service.

An LC record may contain field 010 with a canceled or invalid control number of a previously-distributed record. A record may be canceled because it is a duplicate of the same heading represented by another record. The structure of the canceled/invalid control number is the same as that used by LC in field 001.
(**TYPE 1**)

METS
(Attribute available through the METS Schema on most elements)
ID (ID/O): This attribute uniquely identifies the element within the METS document, and allows the element to be referenced unambiguously from another element or document via an IDREF or an XPTR.
(**TYPE 3**)

(Subelment in METSHEADER)
ALTERNATIVE IDENTIFIERS The alternative identifier element allows one to use alternative record identifier values for the digital object represented by the METS document; the primary record identifier is stored in the OBJID attribute in the root element.
(TYPE 1)

(Attribute of METS root element)

OBJID (string/O): Is the primary identifier assigned to the METS object as a whole. Although this attribute is not required, it is strongly recommended. This identifier is used to tag the entire METS object to external systems, in contrast with the ID identifier.

(**TYPE 1**)

MODS

Element	<recordInfo>
Definition	Information about the metadata record.
Attributes	displayLabel; altRepGroup; lang; xml:lang; script; transliteration
Subelements	<recordContentSource>, <recordCreationDate>, <recordChangeDate>, <recordIdentifier>, <recordOrigin>, <recordInfoNote>, <languageOfCataloging>, <descriptionStandard>
Guidelines	<recordInfo> is a container element that includes subelements relating to information necessary for managing metadata. This type of administrative information can help establish the provenance of a metadata record and may enable better interpretation of the content of the record. <recordInfo> may also include information that is relevant only to the creating or managing institution. <recordInfo> is not repeatable.

Element	<recordIdentifier>
Description	Contains the system control number assigned by the organiztion creating, using, or distributing the record.
Attributes	source; lang; xml:lang; script; transliteration
Subelements	*None*
Guidelines	There should only be on <recordIdentifier> in a record.

(TYPE 1)

@source

Definition
Contains the code or name of the organization whose system control number is located in the <recordIdentifier> element.

Application
The Library of Congress maintains *Organization Code Source Codes* (https://www.loc.gov/standards/sourcelist/organization.html), an online listing of source values. *Organization Code Sources* identifies organization code lists and assigns a code to each. The purpose of this list is to enable the source of organization codes contained in metadata records to be identified by a code.

Element	<identifier>
Description	Contains a unique standard number or code that distinctively identifies a

	resource.
Attributes	type; displayLabel; invalid; altRepGroup; lang; xml:lang; script; transliteration; typeURI
Subelements	*None*
Guidelines	identifier> includes manifestation, expression, and work level identifiers. MODS makes a distinction between a persistent identifier of the resource and its electronic location. This element is used for the former; <location><url> is used for the latter.

(TYPE 2: informational or non-informational)

@type
Definition
Identifies the type of identifier that is recorded.
Application
There is no controlled list of identifier types. Suggested values include, but are not limited to, those specified in *Standard Identifier Source Codes* (http://www.loc.gov/standards/sourcelist/standard-identifier.html)

> Suggested values: hdl, doi, isbn, isrc, ismn, issn, issue number, istc, lccn, local, matrix number, music publisher, music plate, sici, uri, upc, videorecording identifier, stock number

Standard Identifier Sources lists standard number or code systems and assigns a code to each database or publication that defines or contains the identifiers. The purpose of these source codes is to enable the type of standard numbers or codes in metadata records to be indicated by a code.

@typeURI
Definition
Used to indicate that a type is in the form of a URI.
Application
URIs identifying authorities may or may not be dereferenceable to human- or machine-readable information on the authority file, controlled vocabulary, or thesaurus.

Element	<name>
Definition	The name of a person, organization, or event (conference, meeting, etc.) associated in some way with the resource.
Attributes	type; authority; authorityURI; valueURI; usage; displayLabel; nameTItleGroup; altRepGroup; xlink; ID; lang; xml:lang; script; transliteration; etal
Subelements	<namePart>, <nameIdentifier>, <displayForm>, <affiliation>, <role>, <description>, <etal>
Guidelines	<name> is a container element that contains all subelements related to name information. Role values are used to indicate the particular relationship between the name

	and the resource. Some implementors have used the <role> subelement with the value "creator". to retain the concept of main entry. In addition to describing creators, <name> is used as a subelement of <subject>. For names used as subjects, see the <subject> section of the guidelines. A name may be linked to a uniform title in the record using the *nameTitleGroup* attribute. A name may be designated as the citation or "main" entry name using the *usage* attribute.

@type
Definition
Optionally, indicates what type of name is recorded.
Application
The following values may be used with type:

- **personal** – Indicates the name is that of a person.
- **corporate** – Indicates the name is that of a company, institution, or other organization.
- **conference** – Indicates the name is that of a conference or related type of meeting.
- **family** – Indicates the name is that of a family.

@authority
Definition
The controlled list from which the value is taken.
Application
Record the name of the authoritative list used, if applicable, e.g. *authority*="naf". The Library of Congress maintains a list of authority files: Name and Title Authority Source Codes.

@authorityURI
Definition
A URI uniquely identifying the vocabulary from which the controlled term has been selected, as assigned by the body responsible for the maintenance of the vocabulary.
Application
URIs identifying authorities may or may not be dereferenceable to human- or machine-readable information on the authority file, controlled vocabulary, or thesaurus.

@valueURI
Definition
A URI uniquely identifying the term or controlled value from a vocabulary, as assigned by the body responsible for the maintenance of the vocabulary.
Application
URIs identifying terms may or may not be dereferenceable to human- or machine-readable records for the term.

Element	<nameIdentifier>
Description	Contains a unique standard number of code that distinctively identifies the name of a person, organization, or event (conference, meeting, etc.) associated

	in some way with the resource.
Attributes	type; displayLabel; invalid (yes); altRepGroup; lang; xml:lang; script; transliteration; typeURI
Subelements	*None*
Guidelines	This element was introduced in version 3.6 to allow the inclusion of an identifier for the object (RWO) named by this <name>. Repeat the element for each applicable identifier recorded, including invalid and canceled identifiers.

(TYPE 2: informational or non-informational)

TEI

<ident> (identifier) contains an identifier or name for an object of some kind in a formal language.

(TYPE 2: informational or non-informational)

<altIdentifier> (alternative identifier) contains an alternative or former structured identifier used for a manuscript or other object, such as a former catalogue number (structure enables subelements for place and name of agency as well as identification number).

(TYPE 2: informational or non-informational)

<idno> (identifier) supplies any form of identifier used to identify some object, such as a bibliographic item, a person, a title, an organization, etc. in a standardized way

(TYPE 2: informational or non-informational)